In [1]:

```python
import numpy as np
```

In [2]:

```python
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

In [3]:

```python
df=pd.read_csv(r"F:\ML\50_Startups.csv")
```

In [4]:

```
df
```

Out[4]:

	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	Florida	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	Florida	166187.94
5	131876.90	99814.71	362861.36	New York	156991.12
6	134615.46	147198.87	127716.82	California	156122.51
7	130298.13	145530.06	323876.68	Florida	155752.60
8	120542.52	148718.95	311613.29	New York	152211.77
9	123334.88	108679.17	304981.62	California	149759.96
10	101913.08	110594.11	229160.95	Florida	146121.95
11	100671.96	91790.61	249744.55	California	144259.40
12	93863.75	127320.38	249839.44	Florida	141585.52
13	91992.39	135495.07	252664.93	California	134307.35
14	119943.24	156547.42	256512.92	Florida	132602.65
15	114523.61	122616.84	261776.23	New York	129917.04
16	78013.11	121597.55	264346.06	California	126992.93
17	94657.16	145077.58	282574.31	New York	125370.37
18	91749.16	114175.79	294919.57	Florida	124266.90
19	86419.70	153514.11	0.00	New York	122776.86
20	76253.86	113867.30	298664.47	California	118474.03
21	78389.47	153773.43	299737.29	New York	111313.02
22	73994.56	122782.75	303319.26	Florida	110352.25
23	67532.53	105751.03	304768.73	Florida	108733.99
24	77044.01	99281.34	140574.81	New York	108552.04
25	64664.71	139553.16	137962.62	California	107404.34
26	75328.87	144135.98	134050.07	Florida	105733.54
27	72107.60	127864.55	353183.81	New York	105008.31
28	66051.52	182645.56	118148.20	Florida	103282.38
29	65605.48	153032.06	107138.38	New York	101004.64
30	61994.48	115641.28	91131.24	Florida	99937.59
31	61136.38	152701.92	88218.23	New York	97483.56
32	63408.86	129219.61	46085.25	California	97427.84
33	55493.95	103057.49	214634.81	Florida	96778.92

	R&D Spend	Administration	Marketing Spend	State	Profit
34	46426.07	157693.92	210797.67	California	96712.80
35	46014.02	85047.44	205517.64	New York	96479.51
36	28663.76	127056.21	201126.82	Florida	90708.19
37	44069.95	51283.14	197029.42	California	89949.14
38	20229.59	65947.93	185265.10	New York	81229.06
39	38558.51	82982.09	174999.30	California	81005.76
40	28754.33	118546.05	172795.67	California	78239.91
41	27892.92	84710.77	164470.71	Florida	77798.83
42	23640.93	96189.63	148001.11	California	71498.49
43	15505.73	127382.30	35534.17	New York	69758.98
44	22177.74	154806.14	28334.72	California	65200.33
45	1000.23	124153.04	1903.93	New York	64926.08
46	1315.46	115816.21	297114.46	Florida	49490.75
47	0.00	135426.92	0.00	California	42559.73
48	542.05	51743.15	0.00	New York	35673.41
49	0.00	116983.80	45173.06	California	14681.40

In [6]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 5 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   R&D Spend        50 non-null     float64
 1   Administration   50 non-null     float64
 2   Marketing Spend  50 non-null     float64
 3   State            50 non-null     object
 4   Profit           50 non-null     float64
dtypes: float64(4), object(1)
memory usage: 2.1+ KB
```

In [11]:

```
ithubusercontent.com/GaneshSparkz/b5662effbdae8746f7f7d8ed70c42b2d/raw/faf8b1a0d58e251f48a64
```

In [12]:

```
df
```

Out[12]:

	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	Florida	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	Florida	166187.94
5	131876.90	99814.71	362861.36	New York	156991.12
6	134615.46	147198.87	127716.82	California	156122.51
7	130298.13	145530.06	323876.68	Florida	155752.60
8	120542.52	148718.95	311613.29	New York	152211.77
9	123334.88	108679.17	304981.62	California	149759.96
10	101913.08	110594.11	229160.95	Florida	146121.95
11	100671.96	91790.61	249744.55	California	144259.40
12	93863.75	127320.38	249839.44	Florida	141585.52
13	91992.39	135495.07	252664.93	California	134307.35
14	119943.24	156547.42	256512.92	Florida	132602.65
15	114523.61	122616.84	261776.23	New York	129917.04
16	78013.11	121597.55	264346.06	California	126992.93
17	94657.16	145077.58	282574.31	New York	125370.37
18	91749.16	114175.79	294919.57	Florida	124266.90
19	86419.70	153514.11	0.00	New York	122776.86
20	76253.86	113867.30	298664.47	California	118474.03
21	78389.47	153773.43	299737.29	New York	111313.02
22	73994.56	122782.75	303319.26	Florida	110352.25
23	67532.53	105751.03	304768.73	Florida	108733.99
24	77044.01	99281.34	140574.81	New York	108552.04
25	64664.71	139553.16	137962.62	California	107404.34
26	75328.87	144135.98	134050.07	Florida	105733.54
27	72107.60	127864.55	353183.81	New York	105008.31
28	66051.52	182645.56	118148.20	Florida	103282.38
29	65605.48	153032.06	107138.38	New York	101004.64
30	61994.48	115641.28	91131.24	Florida	99937.59
31	61136.38	152701.92	88218.23	New York	97483.56
32	63408.86	129219.61	46085.25	California	97427.84
33	55493.95	103057.49	214634.81	Florida	96778.92

	R&D Spend	Administration	Marketing Spend	State	Profit
34	46426.07	157693.92	210797.67	California	96712.80
35	46014.02	85047.44	205517.64	New York	96479.51
36	28663.76	127056.21	201126.82	Florida	90708.19
37	44069.95	51283.14	197029.42	California	89949.14
38	20229.59	65947.93	185265.10	New York	81229.06
39	38558.51	82982.09	174999.30	California	81005.76
40	28754.33	118546.05	172795.67	California	78239.91
41	27892.92	84710.77	164470.71	Florida	77798.83
42	23640.93	96189.63	148001.11	California	71498.49
43	15505.73	127382.30	35534.17	New York	69758.98
44	22177.74	154806.14	28334.72	California	65200.33
45	1000.23	124153.04	1903.93	New York	64926.08
46	1315.46	115816.21	297114.46	Florida	49490.75
47	0.00	135426.92	0.00	California	42559.73
48	542.05	51743.15	0.00	New York	35673.41
49	0.00	116983.80	45173.06	California	14681.40

In [13]:

```python
df.describe()
```

Out[13]:

	R&D Spend	Administration	Marketing Spend	Profit
count	50.000000	50.000000	50.000000	50.000000
mean	73721.615600	121344.639600	211025.097800	112012.639200
std	45902.256482	28017.802755	122290.310726	40306.180338
min	0.000000	51283.140000	0.000000	14681.400000
25%	39936.370000	103730.875000	129300.132500	90138.902500
50%	73051.080000	122699.795000	212716.240000	107978.190000
75%	101602.800000	144842.180000	299469.085000	139765.977500
max	165349.200000	182645.560000	471784.100000	192261.830000

In [14]:

```python
df["Total_spend"] = df["R&D Spend"] + df["Administration"] + df["Marketing Spend"]
```

In [16]:

```
df
```

Out[16]:

	R&D Spend	Administration	Marketing Spend	State	Profit	Total_spend
0	165349.20	136897.80	471784.10	New York	192261.83	774031.10
1	162597.70	151377.59	443898.53	California	191792.06	757873.82
2	153441.51	101145.55	407934.54	Florida	191050.39	662521.60
3	144372.41	118671.85	383199.62	New York	182901.99	646243.88
4	142107.34	91391.77	366168.42	Florida	166187.94	599667.53
5	131876.90	99814.71	362861.36	New York	156991.12	594552.97
6	134615.46	147198.87	127716.82	California	156122.51	409531.15
7	130298.13	145530.06	323876.68	Florida	155752.60	599704.87
8	120542.52	148718.95	311613.29	New York	152211.77	580874.76
9	123334.88	108679.17	304981.62	California	149759.96	536995.67
10	101913.08	110594.11	229160.95	Florida	146121.95	441668.14
11	100671.96	91790.61	249744.55	California	144259.40	442207.12
12	93863.75	127320.38	249839.44	Florida	141585.52	471023.57
13	91992.39	135495.07	252664.93	California	134307.35	480152.39
14	119943.24	156547.42	256512.92	Florida	132602.65	533003.58
15	114523.61	122616.84	261776.23	New York	129917.04	498916.68
16	78013.11	121597.55	264346.06	California	126992.93	463956.72
17	94657.16	145077.58	282574.31	New York	125370.37	522309.05
18	91749.16	114175.79	294919.57	Florida	124266.90	500844.52
19	86419.70	153514.11	0.00	New York	122776.86	239933.81
20	76253.86	113867.30	298664.47	California	118474.03	488785.63
21	78389.47	153773.43	299737.29	New York	111313.02	531900.19
22	73994.56	122782.75	303319.26	Florida	110352.25	500096.57
23	67532.53	105751.03	304768.73	Florida	108733.99	478052.29
24	77044.01	99281.34	140574.81	New York	108552.04	316900.16
25	64664.71	139553.16	137962.62	California	107404.34	342180.49
26	75328.87	144135.98	134050.07	Florida	105733.54	353514.92
27	72107.60	127864.55	353183.81	New York	105008.31	553155.96
28	66051.52	182645.56	118148.20	Florida	103282.38	366845.28
29	65605.48	153032.06	107138.38	New York	101004.64	325775.92
30	61994.48	115641.28	91131.24	Florida	99937.59	268767.00
31	61136.38	152701.92	88218.23	New York	97483.56	302056.53
32	63408.86	129219.61	46085.25	California	97427.84	238713.72
33	55493.95	103057.49	214634.81	Florida	96778.92	373186.25

	R&D Spend	Administration	Marketing Spend	State	Profit	Total_spend
34	46426.07	157693.92	210797.67	California	96712.80	414917.66
35	46014.02	85047.44	205517.64	New York	96479.51	336579.10
36	28663.76	127056.21	201126.82	Florida	90708.19	356846.79
37	44069.95	51283.14	197029.42	California	89949.14	292382.51
38	20229.59	65947.93	185265.10	New York	81229.06	271442.62
39	38558.51	82982.09	174999.30	California	81005.76	296539.90
40	28754.33	118546.05	172795.67	California	78239.91	320096.05
41	27892.92	84710.77	164470.71	Florida	77798.83	277074.40
42	23640.93	96189.63	148001.11	California	71498.49	267831.67
43	15505.73	127382.30	35534.17	New York	69758.98	178422.20
44	22177.74	154806.14	28334.72	California	65200.33	205318.60
45	1000.23	124153.04	1903.93	New York	64926.08	127057.20
46	1315.46	115816.21	297114.46	Florida	49490.75	414246.13
47	0.00	135426.92	0.00	California	42559.73	135426.92
48	542.05	51743.15	0.00	New York	35673.41	52285.20
49	0.00	116983.80	45173.06	California	14681.40	162156.86

In [17]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 6 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   R&D Spend        50 non-null     float64
 1   Administration   50 non-null     float64
 2   Marketing Spend  50 non-null     float64
 3   State            50 non-null     object
 4   Profit           50 non-null     float64
 5   Total_spend      50 non-null     float64
dtypes: float64(5), object(1)
memory usage: 2.5+ KB
```

In [25]:

```
df_grp_by = df.groupby("State")
```

In [28]:

```python
df_grp_by.sum()
```

Out[28]:

	R&D Spend	Administration	Marketing Spend	Profit	Total_spend
State					
California	1099180.46	2052690.62	3103195.80	1766387.98	6255066.88
Florida	1291584.26	1948302.36	3957176.82	1900384.39	7197063.44
New York	1295316.06	2066239.00	3490882.27	1933859.59	6852437.33

In [33]:

```python
new_yrk_df = df[df["State"] == "California"]
```

In [34]:

```python
new_yrk_df
```

Out[34]:

	R&D Spend	Administration	Marketing Spend	State	Profit	Total_spend
1	162597.70	151377.59	443898.53	California	191792.06	757873.82
6	134615.46	147198.87	127716.82	California	156122.51	409531.15
9	123334.88	108679.17	304981.62	California	149759.96	536995.67
11	100671.96	91790.61	249744.55	California	144259.40	442207.12
13	91992.39	135495.07	252664.93	California	134307.35	480152.39
16	78013.11	121597.55	264346.06	California	126992.93	463956.72
20	76253.86	113867.30	298664.47	California	118474.03	488785.63
25	64664.71	139553.16	137962.62	California	107404.34	342180.49
32	63408.86	129219.61	46085.25	California	97427.84	238713.72
34	46426.07	157693.92	210797.67	California	96712.80	414917.66
37	44069.95	51283.14	197029.42	California	89949.14	292382.51
39	38558.51	82982.09	174999.30	California	81005.76	296539.90
40	28754.33	118546.05	172795.67	California	78239.91	320096.05
42	23640.93	96189.63	148001.11	California	71498.49	267831.67
44	22177.74	154806.14	28334.72	California	65200.33	205318.60
47	0.00	135426.92	0.00	California	42559.73	135426.92
49	0.00	116983.80	45173.06	California	14681.40	162156.86